Error in official age-specific population estimates over place and time

Citation for published version:
Marshall, A, Christison, S & Simpson, L 2017, 'Error in official age-specific population estimates over place and time', *Statistical Journal of the IAOS*, vol. 33, no. 3, pp. 683-700. https://doi.org/10.3233/SJI-160342

Digital Object Identifier (DOI):
10.3233/SJI-160342

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Statistical Journal of the IAOS

Error in official age-specific population estimates over place and time

Alan Marshall[1], Sarah Christison[1], Ludi Simpson[2]

[1] Department of Geography and Sustainable Development, University of St Andrews

[2] Cathie Marsh Institute for Social Research, University of Manchester. Manchester, UK

Keywords: small area population estimates, accuracy, error, area type

Correspondence to:

Dr Alan Marshall

Department of Geography and Sustainable Development

Irvine Building, North Street

University of St Andrews, KY16 9AL

United Kingdom

Tel: +44 (0)1334 463951

E-mail: adm26@st-andrews.ac.uk

Abstract: Population estimates for sub-national areas underpin resource targeting for public and private expenditure. We quantify the success of the Office for National Statistics Small Area Population Estimates (ONS SAPE) in England (2011) using census-based population estimates as a comparative gold standard. We model the accuracy (% absolute error) of the ONS SAPE for Lower Super Output areas according to place characteristics and broad age groups. We compare the modelled accuracy to similar small area population estimates developed by local planners in 1991 and also to simple methods (2011) that might be used with less investment in estimation. Our results show that the ONS SAPE is of comparable accuracy to locally conducted censuses that provided the most accurate results in 1991. We find no combination of area characteristic or age group in which simpler methods of population estimation (in 1991 or 2001) outperform the ONS SAPE. The ONS SAPE is least accurate for young adult ages and areas that are experiencing high unemployment or in-migration. For such areas and groups local censuses may be used to resolve disputes over population estimates and are where attention might be focussed in order to improve the accuracy of small area population estimates.

1. Introduction

1.1 Aims and contribution

Population estimates for sub-national areas underpin resource targeting for public and private expenditure on every service and product, from strategic town planning to mini-markets. They are used directly, as well as informing the denominator of every socio-economic indicator whether death rates, sports participation, or employment. Population estimates that fail to keep up to date with changing population patterns overestimate mortality in an ageing population [1] (p30), and underestimate need for child vaccination in high-fertility areas [2] (p34). Forward planning demands sub-national population forecasts rather than estimates, but the forecasts are impossible without "familiarity with historic trends, and a clear appreciation of emerging directions" as Tom Wilson and Martin Bell [3] put it in a review of small-area demographic modelling (p103).

This study focuses on the accuracy of the products from one of the most successful of governmental population estimation programmes, that of the Office for National Statistics (ONS) for all small areas in England and Wales since 2001 [4]. It now provides accepted annual population estimates for any age group and any area and has almost completely replaced the previous production of estimates in the private and public sectors. The analytical aims of the paper pursued in the following sections are three-fold. First, we aim to quantify the success of the ONS estimates by measuring their accuracy for age groups compared to that achieved before statistical agencies were involved, and compared to simple methods that might be used with less investment in estimation. Second, we investigate the uneven precision of estimates in different types of area, to help producers to identify priority areas of improvement. Finally, we aim to provide users with a means of adding confidence intervals to the estimates, which are sensitive to the type of area each user deals with.

Small areas for the ONS and for this paper are those that cannot be formed by aggregating the standard municipal geography of local authority districts, and are usually smaller than them. The success of the Small Area Population Estimates (SAPE) programme in the 2000s was based on newly developed administrative datasets, on evaluations by the Estimating with Confidence (EwC) research programme of the methods previously used by local government in Britain, and on persistent demand arising from governmental insistence on evidence-based planning for localities during the Labour and Coalition governments 1997-2015 [5].

The remainder of this introductory section briefly reviews methods used for sub-national population estimates in general, and describes the evolution of the method currently used by ONS. It also reviews previous research on the accuracy of population estimates, highlighting the practical results that show that different types of area are associated with different levels of expected accuracy, and that different methods are better at reducing error in specific types of population.

1.2 Methods for small area population estimates

Demographic textbooks tend to ignore the estimation of population when accounts of births, deaths and migration are difficult or impossible to obtain. The international union of demographers' excellent manual of demographic estimation with limited or defective data does not include sub-national population estimation [6]. Small area population estimation is instead a major theme of what has become known as applied demography, with methods reviewed meticulously by Siegel [7] for the United States, and by Rees et al. [1] and Simpson [8] for the UK.

Classification of methods varies between authors. For estimates of the population in a local area, six approaches can be distinguished: mathematical *extrapolation* from more than one past census-based estimate, *apportionment* of an estimate for a larger reference area, *component* methods based on estimates of births, deaths and migration since a census-based estimate, *regression* or *change* methods based on symptomatic indicators such as administrative records or birth counts, local *censuses* involving direct enumeration often riding the back of local electoral enumeration, and finally the *housing unit* method of applying household size to counts of new residential building. Hybrid methods are common, for example an extrapolation of past local time series of censuses, then constrained through apportionment to sum to the estimates for a larger area. The average or other combination of two or more estimates has often been suggested in the academic literature as a means of reducing extreme errors, but it is rarely used in practice [3].

The age-structure for local areas is clearly relevant to services targeted at particular populations such as children, young adults or older adults. In England, sub-national estimates of age structure have been unusually common for many decades because of the availability of records from a universal health service which provide an indicator of the stock of population and an estimate of migration through changes of address. These have become increasingly available for small areas since the 1990s.

EwC was a research partnership between the academic sector and national and local government, around the time of the 1991 Census in the UK. It collated small area population estimates for mid-1991 from over 50 producers for over 5,000 areas, made without the benefit of the 1991 Census results, in order to compare them with each other and with the outcome from the 1991 Census. The project's publications during the 1990s are discussed further below and include the key results of Lunn et al. [9] which are used as a benchmark in this paper.

The EwC evaluations also became one key input to ONS' consideration of their own production of SAPE, which focused on three of the approaches identified by EwC: ratio-change, apportionment, and cohort components [10]. Each would use the increased availability to ONS of individual health records, which since the late 1990s had been collated nationally and used in estimates of age-sex-specific migration for the estimation of larger area populations [11]. ONS work to develop SAPE initially classified areas according to their likely difficulty of estimation, with the aim of establishing which method worked best for each type of area [12]. However, without a longer run of the basic data source of health records for small areas, it was difficult to evaluate the potential of each method, and the more traditional approach of seeking a single method to apply to every area was adopted [10].

1.3 The estimation method used for small areas in England and Wales

The strategy finally chosen by ONS to update mid-year population estimates of small areas since 2001 was ratio-change, and at the time of writing the latest methodological report referred to estimates for mid-2013 [4]. The populations estimated are quinary age groups 0-4 to 85-89, and 90+, in each of the 34 thousand Lower Super Output Areas of England and Wales (population minimum about 1,000 and average about 1,500). The ratio of the most recent year's patient count for the area in each age-sex group to its value the year before is applied to the previous year's population estimate for the same age-sex group. For ages under fifteen the child benefit register was also used to estimate the ratio-change before 2013, when it was discarded because the benefit was no longer universal. Change in the patient register after cleaning and validation is thus assumed to accurately reflect the combined impact of fertility, migration and mortality on each age-sex group. Armed forces and prison populations are estimated separately, from the most recent Census, as they are not included in the national health register. The small area estimates in each of the 348 local authority districts are however adjusted to sum to the estimate for that district at each age-sex group. Therefore, the estimation strategy is a hybrid one, but its main characteristic is ratio-change based on health registers.

These ONS SAPE are now also used to provide estimates for other areas and other age groupings in a very flexible processing procedure. The estimates described in the previous paragraph have gained the authoritative status of National Statistics from the UK Statistics Authority.

The statistical agencies of Scotland and Northern Ireland have also provided population estimates with detail of age and sex for small areas annually since 2001. Scotland uses cohort-component accounts, basing migration on changes of address observed on their health registers, while Northern Ireland uses an average of ratio-change and cohort-component estimates [13].

1.4 Evaluation of the accuracy of small area population estimates

Accuracy is perhaps the main criteria for successful population estimates, though timeliness, cost and disaggregation to geographical, age and other dimensions, can also be decisive. Authoritative literature on evaluation of small area forecasts [14-16, 17] is at least as abundant as it is for estimates [1, 9, 18]. The two concepts are not so different: both are usually constructed by rolling forward the age structure from the previous census, forecasts being taken beyond the current year. The methods of estimation and projection are similar, and evaluative strategies are similar for estimates and for forecasts of small area populations. We supplement the literature on estimates from that on forecasts when it is relevant.

It is clear that the type of population affects the accuracy of estimates. Evaluations repeatedly find that errors are greater with smaller populations, with higher rates of population growth, higher levels of migration in or out, and a longer horizon since the most recent census or other secure estimate. 'Special populations' associated with institutions – prisons, armed forces barracks and educational institutions – are associated with unpredictable population changes that make estimation and forecasting more difficult. Based on this literature, one might argue that population modelling would be more accurate if work were undertaken to find ways of counting special populations more accurately with the administrative data so that they cease to be special.

While there are established results regarding the difficulty of estimating some types of areas, the advantages of one method over another are less clear. Booth's review [19] concludes that "accuracy depends on the particular situation or trends, but it is not clear when a method will perform best" (p547) and "Little progress has been made in advancing knowledge about which methods can be relied upon when conditions are unstable" (p569). This sense that there is an interaction between demographic conditions and the appropriate method was explicit in Rees et al's [1] approach and in their conclusion that ""A comparison of estimation method outputs in GOR East of England has shown that the degree of variation between methods is not substantial for most wards except where special populations are present. However, for groups that may be highly significant for the provision of social services and health needs such as the very young and the elderly, a cohort–component method is needed (unless school pupil, benefits, pension and patient register data are available).” (p30).

In the USA, a study of subnational forecasts Smith and Tayman [20] found that children and young adults were forecast ten and twenty years ahead with most inaccuracy, assumed due to the impact of unpredictable fertility and migration respectively. There was little difference between simple and complex methods of forecasting within a cohort component approach, but the relative accuracy of each method for different types of area or different age groups was not addressed. Rayer and Smith [21] found similar results for counties of Florida, though their age groups 75-84 and 85+ have particularly high percentage error, probably due to the smallness of those elderly populations. Baker et al. [22] examine projections (in spite of the estimates of the title) for the very small Census tracts of Lower California by age and sex, finding high absolute percentage errors. None take a regression approach to their analysis, so that it is difficult to compare the accuracy found in different studies, due to the differing composition of areas considered.

Our search for literature suggest that evaluation of *estimates* of small area population age structure against the outturn of a census or other secure population estimate has remained a neglected research area, since the UK evaluations by the EwC project in the 1990s. These compared for electoral Ward areas a ‘gold standard’ estimate based on the 1991 Census with estimates for 1991 produced independently of the Census. They concurred with the impact on accuracy of size, population growth, levels of migration and special populations found by other studies. For the total population, the gradient of increased error for smaller populations was steep, and a local census was the only method with clearly better accuracy than other methods, going some way to justifying the expense involved [18].

The success of the local census in the EwC project was confirmed in a study of 4,189 population estimates each for five broad age groups, made by 16 different strategies, mainly by local authorities, for 2,008 small areas [9]. Some producers sent more than one set of estimates by different strategies, and the Office of Population, Censuses and Surveys (OPCS), forerunner of ONS, provided a standard they were considering, so that most areas had at least two estimates made for it. This allowed a multilevel regression analysis, which we replicate in this paper, to assess the accuracy of different methods in different types of area. The sixteen strategies were categorised as apportionment, cohort-survival, ratio change, or local census.

The Lunn et al. study [9] found that "for every age group the local census method is at least two per cent more accurate than any other method" (p339). The expected absolute percentage error for the age group 15-24 in an area of average characteristics, for example, was 5% for the local census, 8% for the apportionment methods, 10% for the ratio change methods, and 11% for the cohort-survival method (Fig. 2(i)). The local census accuracy was unaffected by the rate of population change in the preceding decade, though its accuracy was clearly related to the achieved response rate. The cohort-survival methods worked relatively well only with low levels of population change. A key result for the later development of SAPE by ONS was that the one producer using a simple apportionment method that achieved similar accuracy to the local censuses had access to counts of patients from the local health register, as well as information on local numbers of higher education students (p343).

Twenty years on from the EwC programme, we can estimate whether the ONS population estimates for 2011 were more or less successful than their predecessors. In 2007, ONS had declared that "We shall continue to keep our three shortlisted methods under consideration, with a full evaluation planned when the results from the 2011 Census become available," [23] (para 25). Reduced funding meant that the planned evaluation considered only the existing method, and distinguished different levels of accuracy only for different types of administrative area, rather than for the characteristics of areas and populations that are known to affect the accuracy of population estimates. ONS found that "Estimates for Wards, LSOAs, MSOAs, and OAs are less accurate than LAs when compared to 2011 Census, with the degree of difference increasing inversely with the average population size of the area” [24] (p34). Their tabulations for five broad age groups showed that accuracy was better for young and for old age groups. For example, across all Ward areas in England and Wales, the mean absolute percentage error of 4.8% for the population total compares with 4.6% for the age group 0-14, 10.0% for age 15-29, 6.4% for age 30-44, 3.1% for age 45-64, and 3.3% for age 65+.

The rest of this paper takes the evaluation of accuracy of ONS SAPE for LSOAs of England and Wales further in analyses that acknowledge a range of influences on accuracy, and an interaction between the impact of method and of the characteristics of the population. It uses the results to attempt to fulfil the aims of the paper, to highlight the success of ONS SAPE compared to those derived from more simple approaches and previous estimates, the types of area where improvements may be made, and the uncertainty that users should expect in ONS SAPE for specific types of area.

2. Data and Methods

2.1 Population estimates for lower super output areas

The key set of population estimates in this analysis are derived and provided by the Office for National Statistics as part of their standard statistical output, by the method described above. The dataset contains the small area population estimate (SAPE) at mid-2011 for 32,843 lower super output areas (LSOA) according to five age bands (0-14, 15-29, 30-44, 45-64 and 65+) as estimated prior to release of data from the 2011 census.

We have added two population estimates for each LSOA and age group based on simple estimation procedures. These are intended to act as a benchmark for the ONS SAPE. We expect that in general the more sophisticated approach of the ONS SAPE procedure will give more accurate estimates but there may be certain areas where the improvement in accuracy over simple methods is negligible or where a simple approach may actually be more accurate. The two simple population estimates are:

1. **No change**: this set of estimates assume the same age-specific LSOA population count in 2011 as estimated for 2001

2. **Cohort progression**: these estimates involve ageing on the 2001 population by ten years to derive the population in 2011. We assume the population aged 0-14 remains as observed in 2001 (Thus estimates for the 0-14 age group are identical in the No change and Cohort progression estimates).

For each of these simple methods we account for trends in mortality, fertility and migration by calibrating the LSOA (age-specific) population estimates to the ONS mid-year estimate for 2011, rolled forward from 2001 for the local authority district. Thus, like the ONS SAPE, the small area estimates from the simple methods above are adjusted to sum to the estimate for that district at each age group. Our dataset comprises 32,843 areas with 3 estimates in each (ONS SAPE, No change and Cohort progression).

In addition to the methods above we compare modelled error in 2011, across all three methods above, to the modelled error associated with the methods described by Lunn et al. [9]. A full list of methods compared is provided in table 1.

<<<Table 1 about here>>>

2.2 Strategy to measure accuracy of the small area estimates

We follow the EwC strategy in Lunn et al. [9] as far as we are able, in order to make as direct comparisons with the success of the current estimation strategies with those of twenty years before.

Our 'gold standard' to which we compare the ONS and the two crude estimates is the population from 2011 census with a small adjustment reflect the three months between census day (27th March) and mid-year (30th June) (25). In some places we use the word 'truth' as shorthand to describe this census-based population. Unlike in 1991, the 2011 census-based estimates have been widely accepted as the best quality since at least 1981, after a successful effort to gain a coverage that was relatively even between areas [26]. Of course the census-based population estimates are not exact. Occasionally the estimate will be more accurate than the census-based population, but as it is based on an enumeration in 2011, rather than a rolling forward since the enumeration in 2001, it is as close to a gold standard as we can achieve.

2.3 Calibration

Each of our three estimates was calibrated to ONS' rolled forward population estimate for the district that contains the small area as described above. Their quality is therefore due to the methods used for the district estimates as well as to the methods used for the smaller areas. In order to focus *only* on the small area accuracy of the small area methods, we further calibrate all LSOA age-specific estimates to sum to the gold standard (census adjusted mid-year estimate) for each age group in the district. This calibration procedure is in in line with the approach adopted by Lunn et al. [9] enabling a comparison of model results in 1991 and 2011. The calibration is defined below where:

P_{ijak} is the mid-2011 rolled forward population estimate for LSOA i and method j at age a within district k

P_{ak} is the mid-2011 rolled forward population estimate for age a and district k

C_{ak} is the census adjusted mid-year population estimate for age a and district k.

Then O_{ijak} , the calibrated mid-2011 population estimate for LSOA i, method j, age a within district k is defined as:

$$\boldsymbol{O_{ijak}} = \frac{\boldsymbol{C_{ak}}}{P_{ak}} \times \boldsymbol{p_{ijak}}$$

For the ONS SAPE, The impact of the calibration is to decrease the mean absolute percentage error as illustrated in table 2. Thus, on average LSOA SAPE are moved towards LSOA census estimates through calibration to district census estimates. As a result of this calibration, we observe zero bias in the SAPE estimates when compared to LSOA census adjusted SAPE.

<<<Table 2 about here>>>

We follow most other evaluations in measuring the mean absolute percentage error: "In terms of what is important for producers and users of population estimates, it seems that the absolute percentage inaccuracy is the most suitable candidate for analysis, since it reflects the uncertainty associated with estimates" [9] (p334). For each population estimate (ONS SAPE, No Change and Cohort Progression) we calculate E_{ij} where i indexes the 32,843 LSOAs (i=1,…….,32,843) and j indicates the multiple estimates within areas according to method (j=1,2,3). For simplicity we drop the age a and district k subscripts from the algebraic specifications below.

$$E_{ij} = \left(\frac{|O_{ij} - C_{ij}|}{C_{ij}}\right) * 100$$

And:

O_{ij} = Mid-2011 estimate for LSOA i and method j (calibrated so that LSOAs sum to the Census-based gold standard for the wider district area)

C_{ij} = Mid-2011 Census-based population for LSOA i and method j

Our dependent variable is a log transform of E_{ij} (absolute % error in area i and method j), using a constant that eliminates the skewed distribution of E_{ij} (see figure 1):

$$z_{ij} = ln(E_{ij} + 1.4)$$

<<<Figure 1 about here>>>

Figure 2 shows the distribution of the error for each of the three estimates, with an error much lower on average in the ONS SAPE (mean absolute % error = 7.1%) compared to that in the No change (mean absolute % error = 16.2%) and Cohort progression methods (mean absolute % error = 17.6%). There is a considerable overlap in the range of error across the three methods, so it is possible for some areas (and potentially areas with particular characteristics) to be estimated more accurately using one of the two simpler methodologies rather than the ONS SAPE.

<<<Figure 2 about here>>>

We compare our results on the accuracy of SAPE to the analysis of Lunn et al. [9], considering whether there have been improvements in level of error and whether we observe the same relationships between area characteristics and error. The main differences are as follows. Our five age groups, taken from material published alongside the associated ONS report [24] are: 0-14, 15-29, 30-44, 45-64 and 65+. Lunn et al. [9] also used five age groups, and they are the same except for the age-group boundary at 29: the 1991 study used 15-24 and 25-44. We use all LSOAs in England and Wales, which have a relatively homogenous total population, while Lunn et al. [9] examined about a third of electoral wards, which have a more widely varying population which is on average larger than LSOAs. The analysis of Lunn et al. [9] involved 7,627 population estimates across 4,383 areas and so, unlike our analysis, not all methods were represented in each area. Our analytical regression approach is as similar as we have been able to replicate, and takes account of population size, so that we can compare the accuracy in 1991 and 2011 that would be expected for areas of the same size as well as for other comparable characteristics. There are some small potential improvements on the modelling that we do not employ in order to preserve comparability, in particular the use of quadratic term for some of the independent variables. We have left the accuracy of the estimate of total population to separate analyses as in the EwC studies [18], but do not expect the relationships between method and area characteristics to be different from those found here. In presenting our results we compare modelled error from the various 1991 estimates (derived from our own calculations using regression coefficients as stated in Lunn et al. [9] with the 2011 estimates' modelled error (derived from the regression coefficients of the final model described below). Predicted errors relate to matched values of independent variables wherever possible (the same value for population size, population growth, unemployment, institutional population, in-migration). We were not able to include the percentages of students or armed forces in our data of area characteristics for 2011, which as we discuss may create some differences in other results.

2.4 Independent variables

The inclusion of specific explanatory variables is informed by the discussion of literature above, and the ability to compare with Lunn et al. [9]. We expect estimates to be less accurate according to an area's migration, the size of the population and the extent of population growth, the % of populations living within institutions and the level of unemployment as a marker of deprivation.

All the area data are taken from the 2011 census except for the 2001 estimates of population used to derive population growth. We convert 2001 population estimates from 2001 LSOA boundaries to 2011 LSOA boundaries using a geographical conversion table derived from the Geoconvert website. The distribution of the area variables is shown in figure 3.

<<<Figure 3 about here>>>

Table 3 gives descriptive statistics for the dependent variable before transformation in 1991 and 2011 and the independent variables. Some clear patterns of different precision between the methods emerge. However, as the 1991 and 2011 evaluations used different sets of areas, we cannot draw conclusions from the crude differences shown in Table 3 without first accounting for the nature of the areas, which we model as now described.

<<<Table 3 about here>>>

2.5 Model

We use a multilevel model, with our estimates of precision across the three methods, nested within areas. Such a model is appropriate given the expectation that the precision of multiple estimates within the same area will be correlated; for example, in some areas the process of population estimation is likely to be more challenging (e.g. if populations are highly mobile or less well represented in administrative data) and so the precision of estimates from all models will be low. We include random coefficients for the intercept and for each of the method dummy variables with the expectation that error, and the difference in error according to method, will vary between areas.

We test interactions between method and each area characteristic and between method and each age group. Thus, we investigate whether the relationship between method and error varies according to age and according to characteristics of place. By including such interactions we can assess whether particular methods are better/worse in terms of error over various combinations of area types and age groups. We centre all the area independent variables in our model by subtracting the overall mean value of each area characteristic to ease interpretation of model results. As a result the constant gives the mean error for an LSOA that takes the average value for each of the area independent variables. The model specification is given below:

$$z_{ij} = \beta_{0i} + \sum_{k=2}^{5} \beta_k x_k + \sum_{j=2}^{3} \delta_{ji} m_j + \sum_{l=1}^{5} \gamma_l f_{lk} + \sum_{j=2}^{3}\sum_{k=2}^{5} \alpha_{jk} m_j x_k + \sum_{j=2}^{3}\sum_{l=1}^{5} \rho_{jl} m_j f_{lk} + e_{ij}$$

With

$\beta_{0i} = \beta_0 + U_{0i}$

$\delta_{2i} = \delta_2 + U_{2i}$

$\delta_{3i} = \delta_3 + U_{3i}$

And

$$\begin{pmatrix} U_{0i} \\ U_{2i} \\ U_{3i} \end{pmatrix} \sim N(0, \Omega_u)$$

$$\Omega_u = \begin{pmatrix} Var(U_{0i}) & Cov(U_{0i,}U_{2i}) & Cov(U_{0i}, U_{3i}) \\ Cov(U_{2i}, U_{0i)} & Var(U_{2i}) & Cov(U_{2i}, U_{3i} \\ Cov(U_{3i}, U_{0i} & Cov(U_{3i}U_{2i}) & Var(U_{3i)} \end{pmatrix}$$

$e_{ij} \sim N(0, \sigma_2^2)$

Where:

x_k=1 if age=k and 0 otherwise (k=1,….,5; 0-14, 15-29, 30-44, 45-64, 65+)

m_j=1 if method =j and 0 otherwise (j=1,2,3; ONS SAPE, No change, Cohort progression)

f_{lk}= the value of the l (l=1,…,5) area (LSOA) variables including 2011 population size, % institutional population, % unemployment, % population change between 2001 and 2011 and %in-migration, of which two (% population size and % population change) distinguish detail of k (k=1,…,5) age groups

2.6 Outliers

Examination of regression diagnostics, and the skewed nature of some of the independent variables on area characteristics, may be problematic for model fit. In order to test this possibility we replicated our reported analysis excluding cases with a Cook's distance greater than 5 (2 times the mean Cook's distance). The results presented in the paper include all LSOAs, but the substantive findings reported also hold for the analysis excluding outliers based on Cook's distance (results from the outlier sensitivity testing are available on request from the authors).

3. Results

3.1 Summary of key results

The results in Tables 4 and 5 and Figure 4 demonstrate five key findings. First, the ONS SAPE outperform other simpler methods in 2011 across all areas types although the differential in error is smaller for those areas with characteristics associated with the greatest error (for example, high unemployment and high in-migration). Second, the ONS SAPE offers a clear improvement in accuracy over each of the non-census estimates in 1991 across each age group and the range of area characteristics included the model. Third, the ONS SAPE provides a comparable level of accuracy to the local census population estimates, the most accurate method in 1991. Fourth, the direction (and magnitude) of the associations between error and area characteristics for the 2011 ONS SAPE and the

1991 census are broadly similar, although we do see some evidence for a steeper increase in error with rising area unemployment for the 2011 SAPE compared to the 1991 census. Finally, the average size of error at the 0 to 14 age group that might be expected by users of the ONS SAPE (ten years after a census) decreases from 7.7% to 3.9% as one moves from local areas of low population, high unemployment and high in-migration to areas of high LSOA population, low unemployment and low in-migration.

3.2 Distribution of error between small areas and within small areas

Table 4 gives the estimated parameters from the regression model of the log transformed absolute % error in 2011 (after calibration to the district 'truth'). In model 1, the variance components model, we see that 10% of the variability in error across LSOAs is attributable to areas. However, after controlling for area characteristics, method and age (model 2), the variability in error attributable to area drops to 8%. The area characteristics we include in our model explain around half the variability in error between areas and some unmeasured characteristics of LSOAs continue to influence error in our full model. The remaining random error with variance of 0.53 within small areas and 0.03 between them is very close to that found in the analyses conducted for 1991: 0.49 within areas and 0.06 between areas. Most variation between the estimates' error that we have not been able to account for is due to different estimates for the same area, rather than the differences between areas. The total residual variance of 0.56 in the transformed dependent variable, with constant 1.85, translates to a distribution for the untransformed percentage error that has a 95% confidence interval of 27% points. This random error is mainly within areas. Between areas, the remaining residual variance implies an error with 95% confidence interval of width 4.6%. This puts in context the magnitude of errors that we find between methods and between types of area.

<<<Table 4 and Figure 4 about here>>>

In model 3 we extend model 2 to allow each of the Method dummy variables to vary randomly across places. The random effect associated with each method dummy variable (No change and Cohort progression) indicate that the higher error for each of these simpler methods does indeed vary across LSOAs even after controlling for the other explanatory variables. For example, from table 4, the 'No change' model results in an increase in the log transformed error of 0.73 (fixed effect), but across areas the random effect indicates that the differential varies between 0.73+1.96*0.03=0.79 and 0.73-1.96*0.03=0.69.

Figure 4 shows the change in predicted absolute % error across the three methodologies tested in 2011 and the four tested in 1991, according to age group and the area characteristics included in the model. All predicted values for 2011 relate to the model 3 in table 4. Figure 4a shows that the crude methods we constructed for 2011, using only the 2001 age structure and calibration to ONS district SAPE, are worse than the 2011 SAPE and the 1991 local census. The ONS SAPE has an improvement over these crude methods of more than 3 percentage points at each age group, nearly halving their inaccuracy from 6-11% to 4-7%, depending on the age group. However, in some combinations of area characteristic and age group the simple 2011 approaches are comparable or more accurate than the 1991 non-census methods. For example, the No change and Cohort progression methods regularly have lower % absolute error compared to the Ratio method in 1991. Each method including the ONS SAPE has most difficulty estimating the number of young adults. Errors for the age group 15-29 should be expected to be about double those for children or for adults 45-64 or 65+.

A striking result from this comparison of error and its correlates over time is that the 2011 ONS achieved approximately equal accuracy compared to local censuses, the most accurate of the 1991 SAPE methodologies The exception to this is for difficult to estimate areas (e.g. high unemployment) and the challenging age group of young adults. For this young adult age group we see that the local census is more accurate than the ONS SAPE by 1 percentage points. For areas with unemployment at 35%, the local censuses gave more accurate estimates by around 2 percentage points. Conversely, for areas of large population (around 1000), the ONS SAPE is, on average, more accurate than local censuses by 1 percentage point. There is little difference between the accuracy of the ONS SAPE and the 1991 local censuses across the area distribution of in-migration, institutional populations and population change.

3.3 Effects of method, age group and area characteristics on error

In line with the literature we see that the accuracy of population estimates improve as the population size of the LSOA increases for all methods (figure 4b.). For the ONS SAPE, there is a 1 percentage point improvement in accuracy of estimates moving from a population size of to 200 to 1000. Error and population change (figure 4c.) are, not surprisingly, strongly positively associated for the most simple methods (No change and Cohort progression) which do not effectively accommodate such population change. However, LSOA population growth has relatively little effect on the accuracy of estimates from the ONS SAPE, less so than the estimates from 1991, with the exception of the local census. It seems that the health administrative records capture population change for different age groups better than the records that existed before 1991.

As the % of in-migration increases the error in SAPE for all 2011 methods increases (figure 4d); for the ONS SAPE there is an increase in error of 3 percentage points (3% to 6%) as the % in-migration increases from 0% to 30%. For the 1991 estimates we see a drop in error with higher in-migration for apportionment and ratio methods, so that in areas of high migration these methods give more accurate estimates, however, the this negative association was not statistically different from zero in the modelling.

The relationship between absolute % error and % unemployment (figure 4e.) is characterised by increasing error with increasing unemployment across all methods. The error of the ONS SAPE increases from around 4% to 7% as the level of LSOA unemployment rises from 10% to 30%. Interestingly, the gradient of the association between area unemployment and error is shallower for all the other estimation methodologies compared to ONS SAPE so that the improved accuracy of the ONS SAPE is less in areas experiencing high unemployment. For example, in areas with high unemployment rates of 25%, the ONS SAPE method is more accurate than the 1991 Apportionment method by around 2.5 percentage point compared to a difference of 4 percentage points at 5% unemployment.

For all three methods for 2011 we observe very little change in % error with increasing levels of institutional population (figure 4e.), similar to the flat relationship found for many of the 1991 estimates.

3.4 Summary of error in ONS SAPE

In general, the ONS SAPE have a lower level of error than the 1991 non-census estimates, with the exception of the local census. For example, we observe at least 3% lower error in the 2011 SAPE compared to the best 1991 non-census estimate across all population sizes. There only exception to this general finding occurs for areas experiencing high in-migration where estimates from the ONS SAPE and 1991 Cohort survival have similar levels of error. At 30% in-migration we observe an error of 7% for the 2011 SAPE and 1991 Cohort survival. Finally, the error in estimates is lower across all levels of institutional population for ONS SAPE compared to the 1991 modelled error. Thinking of ONS SAPE in particular, it is probable that administrative health records are likely to include residents in institutions and therefore monitor change well. It is one difference between the performance of population estimates where available updated information can track precipitous changes, and the performance of population projections where such changes are unpredictable and major sources of error.

The direction of association between area characteristics and error is strikingly similar for the ONS SAPE and the 1991 local census. For both methods error is; positively correlated with in-migration and unemployment; negatively associated with population size and exhibits a flat association with population change and % in institutional populations. There is a suggestion that the magnitude of association between error and unemployment and between error and population size appears greater for the ONS SAPE than the 1991 census. Another appropriate comparison is between the 1991 ratio method and the ONS SAPE which also used the ratio approach. Again, the two methods have the same direction of association between most of the area characteristics and error, albeit with different gradients of association. In both 1991 and 2011 the ratio method generates error in its SAPE that is positively correlated with % in-migration (stronger in correlation in 2011), % unemployment (stronger correlation in 2011) and % population change (stronger correlation in 1991) and negatively correlated with population size (similar strength of association in 1991 and 2011).

3.5 Magnitude of error that might be expected by users of the ONS SAPE

All of the results so far have examined modelled error for small area population counts after the counts are calibrated to sum to the 'truth' for wider district areas for reasons noted in the calibration sub-section of the Data and Methods section. These calibrated estimates are valuable for methodological reasons and for comparison to the analysis of Lunn et al [9], but the calibrated error is generally lower than from uncalibrated estimates (see table 2). Table 5 shows the expected % absolute error in the ONS SAPE that might in practice be encountered by a user of these statistics across the distribution of population size, % unemployment and % in-migration. Here we do not calibrate estimates to the 2011 district truth (census adjusted mid-year population estimates for districts) as we wish to describe the predicted error that would be faced by a user of these statistics which do not contain such calibration prior to the release of census data. The model fitted is identical to that in table 4 (model 3) except the dependent variable is not calibrated to 2011 census adjusted mid-year population estimates for districts. We use a similar log transform as described earlier but require a constant of 1.5 rather than 1.4. Analysis of table 5 suggests that an increase in population from the 10th percentile to the 90th percentile is associated with an absolute decline in error of around 0.6% across the distribution of area unemployment or in-migration. An increase in error of around 2%

is observed as one moves from the 10^{th} percentile of to the 90^{th} percentile of area in-migration. Similarly, we observe an increase in error of just over 1% moving from the 10^{th} percentile to the 90^{th} percentile of area unemployment. Considering extremes of error, users might expect an increase in average error from 3.9% to 7.7% as one moves from an area with high population (75^{th} percentile), low unemployment (10^{th} percentile) and low in-migration (10^{th} percentile) to an area of low population (25^{th} percentile), high unemployment (90^{th} percentile) and high in-migration (90^{th} percentile).

<<<Table 5 about here>>>

4. Discussion

Before turning to answer the three questions that motivated this study, an acknowledgement of two weaknesses is in order, though we believe they do not threaten our conclusions. First, any differences in error between 1991 and 2011 may be a result of changes in the challenge of estimating populations at two time points rather than differences stemming from the particular method employed. It may be that people have become harder to monitor between censuses, in spite of the greater existence and use of pooled administrative data. The success of the local censuses in 1991 with relatively high response rates may understate the achievement of the ONS SAPE in 2011 or later years, for example. Second, we have treated the 2011 census as the ‘truth’ by which we evaluate each estimate, but the census is also prone to error which is strongly patterned by age and area characteristics. We think that the effort involved in the census and its adjustment for non-response suggests that the census-based mid-year estimate in the year of the census is *better* than any of our estimates as a measure of the true population. However, error in the census-based ‘truth’ makes it wise to treat the differences that we have found between it and other estimates of population as an upper bound of their error at that time, as it includes an element of the error in the census itself.

The first question we set out to answer was the extent of the success of ONS (2011) SAPE compared to those derived from more simple approaches and previous estimates. The results section has demonstrated that area characteristics and the age group of population under estimation are associated with error in ways that are largely consistent in 1991 and 2011 and consistent with previous literature: it is most difficult to estimate the population of young adults, and of areas with smaller populations, populations that are changing rapidly, or which are in relative poverty (as measured here by unemployment) or which have high turnover of population. The consistent pattern of error across ages are likely to be a result of the age-specific intensity of migration and the challenges of estimating migration. Young adults are known to be most mobile and so it is not surprising that this age group is less well estimated in both 1991 and 2011 and across methods. Similarly, areas with small populations are likely to experience greater changes in population, relative to their population size, over short time periods that are challenging to estimate and thus the association between population size and error in 1991 and 2011 is unsurprising.

That the ONS SAPE in 2011 achieves comparable accuracy compared to local censuses in 1991 is a major achievement, with considerable savings in terms of cost and effort. However, our results provide some evidence to suggest that local censuses remain the most reliable way of estimating the populations of small areas particularly for hard to estimate areas and age groups, and thus remain a sensible option to settle disputed population estimates. The ONS SAPE is more accurate than the other non-census methods. This success of the ONS SAPE is reinforced by its good performance for children and older people, and its relative immunity to increased error in fast-changing populations. However, the extent to which it provides more accurate estimates than non-census methods in 1991 is reduced for other populations which challenge estimates generally: young adult ages and areas that are experiencing high unemployment or high in-migration.

Our comparison of area correlates of error in 1991 and 2011, and the stronger associations between error and % unemployment and % in-migration in 2011, suggest that urban, young and deprived populations are becoming harder to count as has local accounts of migration as a component of area demographic change. Why should such change have occurred? It is not the case that internal migration has become more common over this period and so this is unlikely to be an important driver of the higher error observed in areas experiencing high levels of migration. An alternative explanation is that poorer quality administrative data may reflect a similar processes of declining willingness to participate in social and civic areas of society as evidenced by declines in the proportion of adults registered to vote for example.

Where might improvements be made? Within the current strategy of monitoring the local rate of change in patient registers for each age and sex, the results focus attention on the quality of those registers for young adults and in areas that are poorer or have high population turnover, where the ONS SAPE was outperformed by 1991 local census estimates. It is no surprise that these populations will be less easily represented in an up-to-date manner by health records. It is of concern that estimation of the child population, which was notably well-estimated by the ONS SAPE in 2011, has since been deprived of one of its two contributing data sources, universal child benefits. Other administrative data are being actively considered, and we return to this development at the end of this discussion, but the change in an administrative data source is a warning about the difficulty of establishing a reliable estimate of population estimates based upon them.

What level of uncertainty should users expect in ONS SAPE for specific types of area? We have used our results to provide Table 5 which shows that estimates of population aged 0-14 should expect to carry around 5% error – either over-estimate or under-estimate, with about half of estimates more than this and half of them less. These are errors expected for estimates ten years after the previous census. Without other evidence one might assume for practical purposes that the error increases linearly with years after the previous census. Table 5 allows users of estimates to read off the impact of population size, unemployment and in-migration, all of which are characteristics that users are likely to have estimates of for each local area. The range of expected error between the easiest areas to the most difficult (small to large populations, low to high unemployment, and low to high migration) is from about 4% to about 8%.

Clearly a single table cannot encompass all the advice that can be given to users. Our modelling approach however lends itself to presentation of expected errors in the more flexible format of spreadsheets with adjustable inputs for area characteristics, or a list of all standard areas with expected errors and confidence intervals for each age groups and according to each area's known characteristics from latest census information.

There is plenty of scope for further evaluation in addition to presentation of practical use to consumers of the population estimates. Modelling of the error of the total population, and of areas other than LSOAs is a priority. Extension to the methods used in Scotland and Northern Ireland would be particularly instructive because cohort component accounts inform their small area population estimates. ONS are currently exploring other means of population estimation as part of an attempt to augment or replace the decennial census, and have published experimental estimates of the population by age and sex, for LSOAs in England at mid-years from 2011 to 2015 [27]. These are based on the active records from not only health registers but tax, student, school-pupil, welfare benefits and other databases. To replace the decennial census and gain the support of applied demographer and others who depend on sub-national population estimates, they will need to have proven accuracy considerably greater than the current estimates ten years after the previous census that have been evaluated here. With this variety of approaches, it is also feasible to evaluate the potential of estimation approaches specific to types of area, and the use of composite approaches that pool population estimates across methods. The regression approach to evaluation of small area population estimates is capable of identifying the different contributions to accuracy of method, type of population, and characteristic of areas.

Acknowledgements

We are grateful to Steve Smallwood, Becky Tinsley and others at the Office for National Statistics for the very helpful comments and feedback that they provided on this paper and for providing the 2011 data that underpinned the analysis.

References

1. Rees P., Norman P., Brown D. A framework for progressively improving small area population estimates. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 2004; 167(1): 5–36.

2. Cavenaghi, S., Ed.. Estimaciones y proyecciones de poblacion en America Latina: *Desafios de una agenda pendiente. Serie e-Investigaciones* 2. Rio de Janeiro, Asociacion Latinoamericana de Poblacion. 2012.

3. Wilson T., Bell M. Editorial : Advances in local and small-area demographic modelling. *Journal of Population Research* 2011; 28(2/3): 103–107.

4. ONS. Methodology Note on production of Small Area Population Estimates. Titchfield, Office of National Statistics. 2015a.

5. DCLG. A plain English guide to the Localism Act. London. Department of Communities and Local Government. 2011

6. Moultrie, T., R. Dorrington, et al. Tools for demographic estimation. Paris; IUSSP. 2013

7. Siegel J.S. Applied demography: applications to business, government, law, and public policy. San Diego: Academic Press.; 2002.

8. Simpson S. Making Local Population Statistics: a Guide for Practitioners for the Estimating with Confidence Project. Wokingham:. Local Authorities Research and Intelligence Association. 1998

9. Lunn D, Simpson, S, Diamond, I, Middleton, L. The accuracy of age-specific population estimates for small areas in Britain. *Population Studies* 1998; 52(3): 327-344.

10. Bates A. Methodology used for producing ONS's Small Area Population Estimates, *Population Trends*,125: 30-36. 2006.

11. Scott A, Kilbey T. Can Patient Registers give an improved measure of internal migration in England and Wales? *Population Trends*,96: 44-56. 1999.

12. ONS. Categorisation of areas, Working Group on small area population estimates, SAPEWG (03)2. Titchfield, Office for National Statistics. 2003.

13. NISRA. Small Area Population Estimates across the UK. Belfast, Northern Ireland Statistics and Research Agency. 2011.

14. Ahlburg D.A. Simple versus complex models: evaluation, accuracy and combining. *Mathematical Population Studies* 1995; 5(3): 281–290.

15. Smith S.K, Shahidullah M. Evaluating population projection errors for census tracts. *Journal of the American Statistical Association* 1995; 90: 64-71.

16. Rayer S, Smith S. K. Factors Affecting the Accuracy of Subcounty Population Forecasts. *Journal of Planning Education and Research* 2010; *30*(2), 147–161.

17. Wilson T. New evaluations of simple models for small area population forecasts. *Population, Space and Place* 2015*;* 21(4): 335-353.

18. Simpson S, Diamond I, et al. Updating small area population estimates in England and Wales. *Royal Statistical Society A* 1996; 159(2): 235-247.

19. Booth H. Demographic forecasting: 1980 to 2005 in review. *International Journal of Forecasting* 2006; 22(3), 547–581.

20. Smith S.K, Tayman J. An Evaluation of Population Projections by Age. *Demography* 2003; 40(4): 741-757

21. Rayer S, Smith S.K. Population projections for age for Florida and its counties: Assessing accuracy and the impact of adjustments. *Population Research and Policy Review* 2014; 33(5): 747-770.

22. Baker J, Alcantara A , Ruan X, Watkins K, Vasan S. A comparative evaluation of error and bias in census tract-level age/sex-specific population estimates: Component I (Net-Migration) vs Component III (Hamilton–Perry) *Population Research Policy Review* 2013; 32(6): 919–942.

23. ONS. Update on the small area population estimates project, Steering group on small area population estimates Paper SAPE(07)1. Titchfield, Office for National Statistics. 2007.

24. ONS. Small Area Population Estimates (SAPE) Evaluation: Report on Accuracy Compared to Results of the 2011 Census. Titchfield, Office for National Statistics. 2015b.

25. ONS. Methodology Note on production of Super Output Area Population Estimates. Titchfield, Office for National Statistics. 2013.

26. ONS. Data Quality. Chapter 8, pp 215-227 in ONS, 2011 Census: General report for England and Wales. Titchfield: Office for National Statistics. 2015c.

27. ONS. Research outputs estimating the size of the population in England and Wales: 2016 release. Titchfield: Office for National Statistics. 2016.

Figures

Error	Log (absolute % error+1.4)

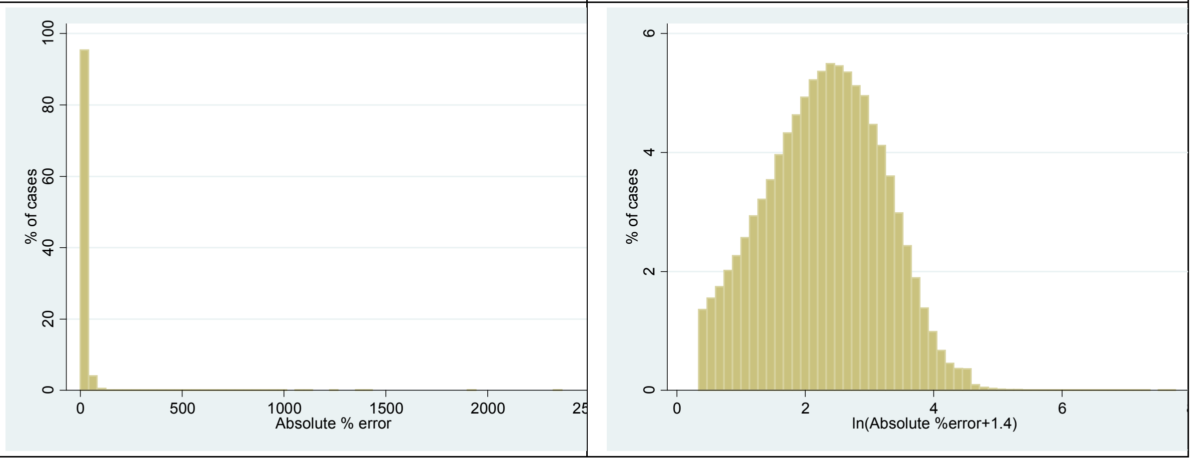


Figure 1: Distribution of absolute % error and its log transform

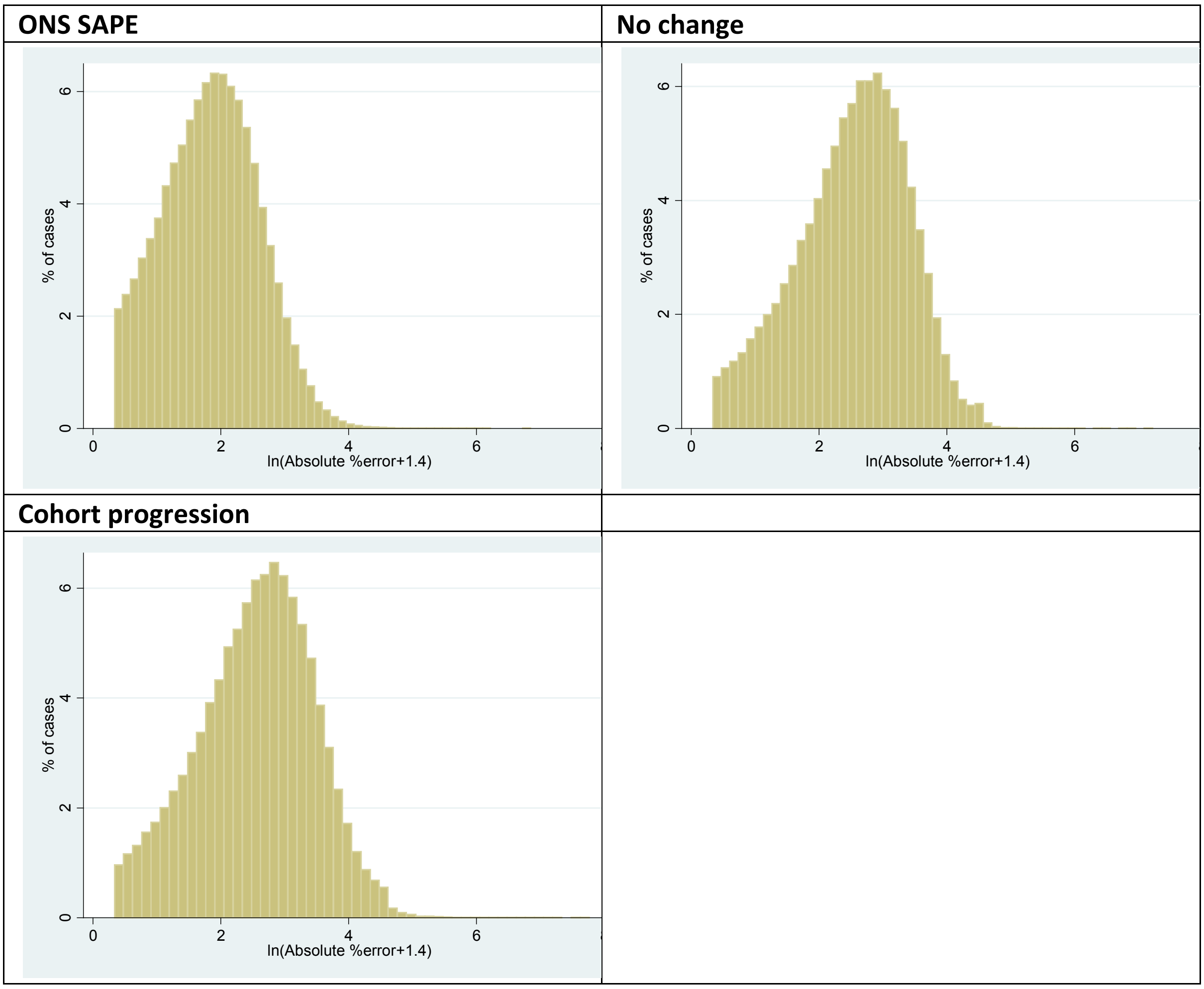


Figure 2: Distribution of the dependent variable, ln(absolute % error+1.4), across each SAPE method (ONS, No change, Cohort progression)

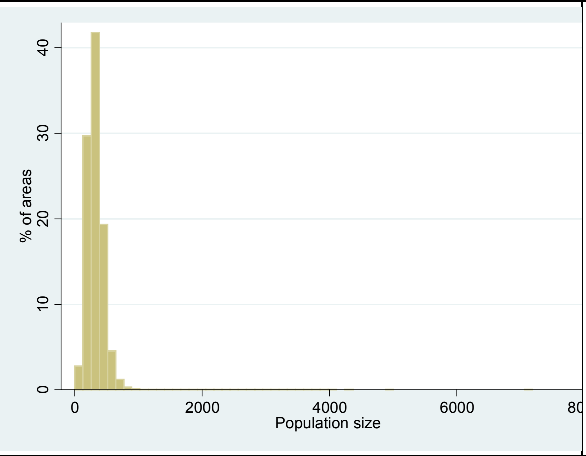
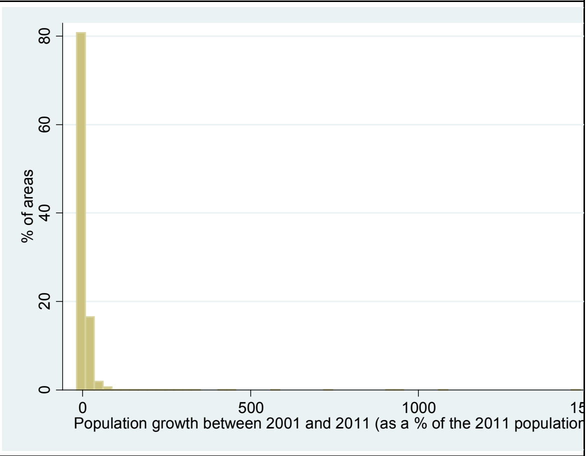
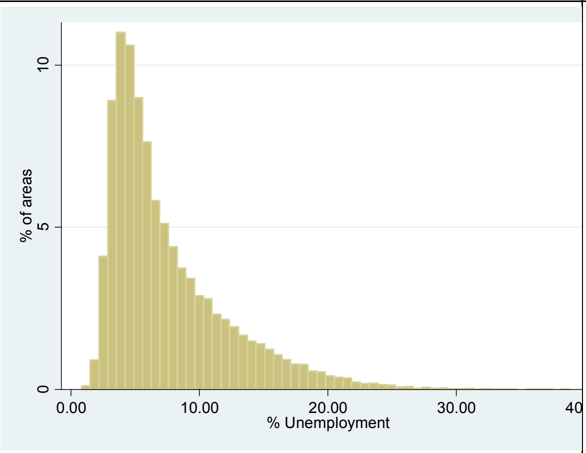
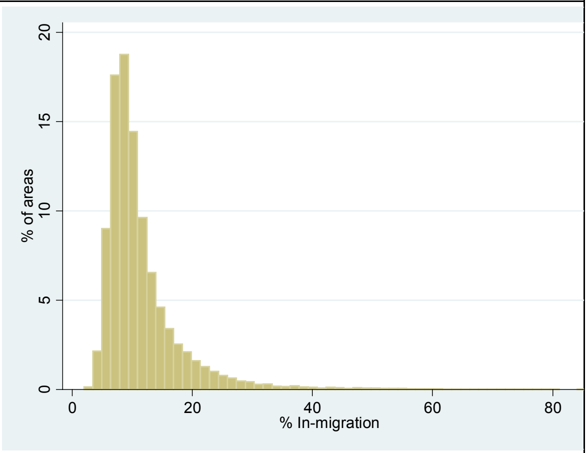
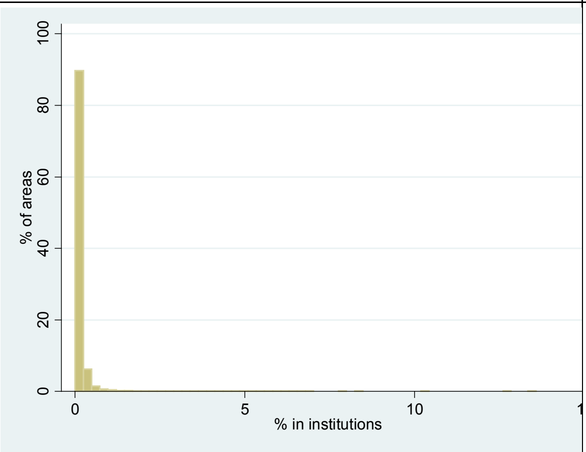
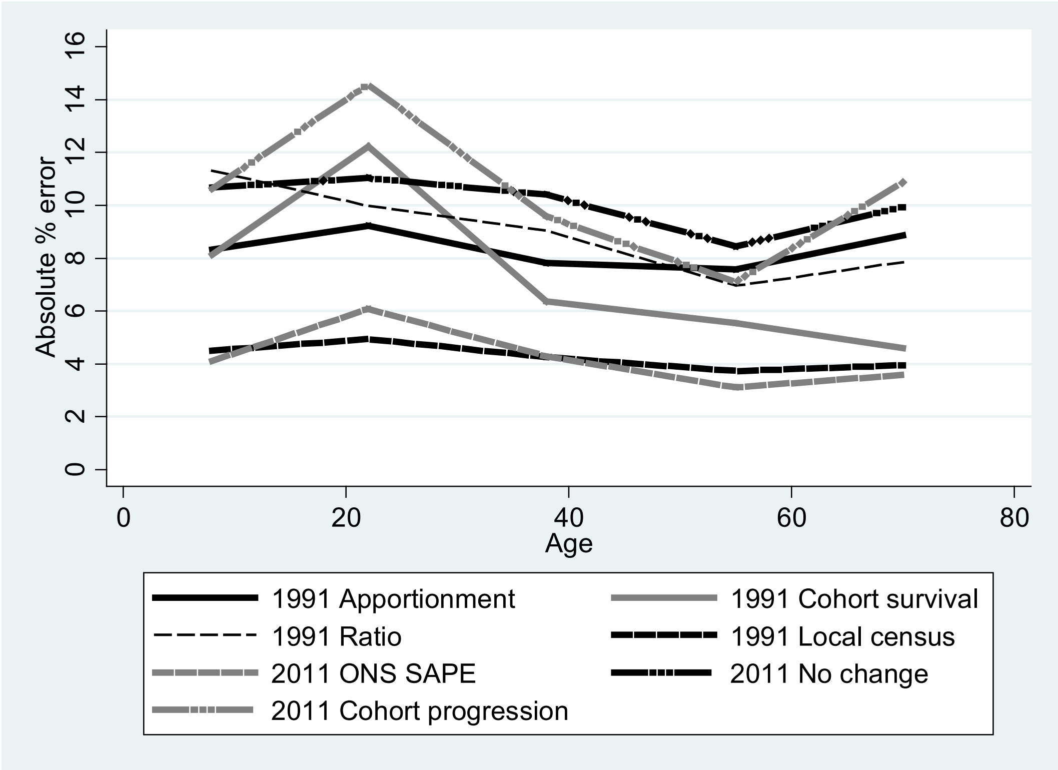
Distribution of LSOA Population size (age specific)	Population growth
	
Distribution of % unemployment across LSOAs	Distribution of % in-migration across LSOAs
	
Distribution of % institutional population across LSOAs	
	

Figure 3: Distribution of explanatory area variables (note: some variables are clearly skewed with evidence of outliers. All the conclusions and results hold after excluding cases with Cooks distance greater than 2*mean cooks distance)

4a. Model absolute % error by age group and method (1991and 2011)



4b. Model absolute % error by population size and method (1991and 2011)

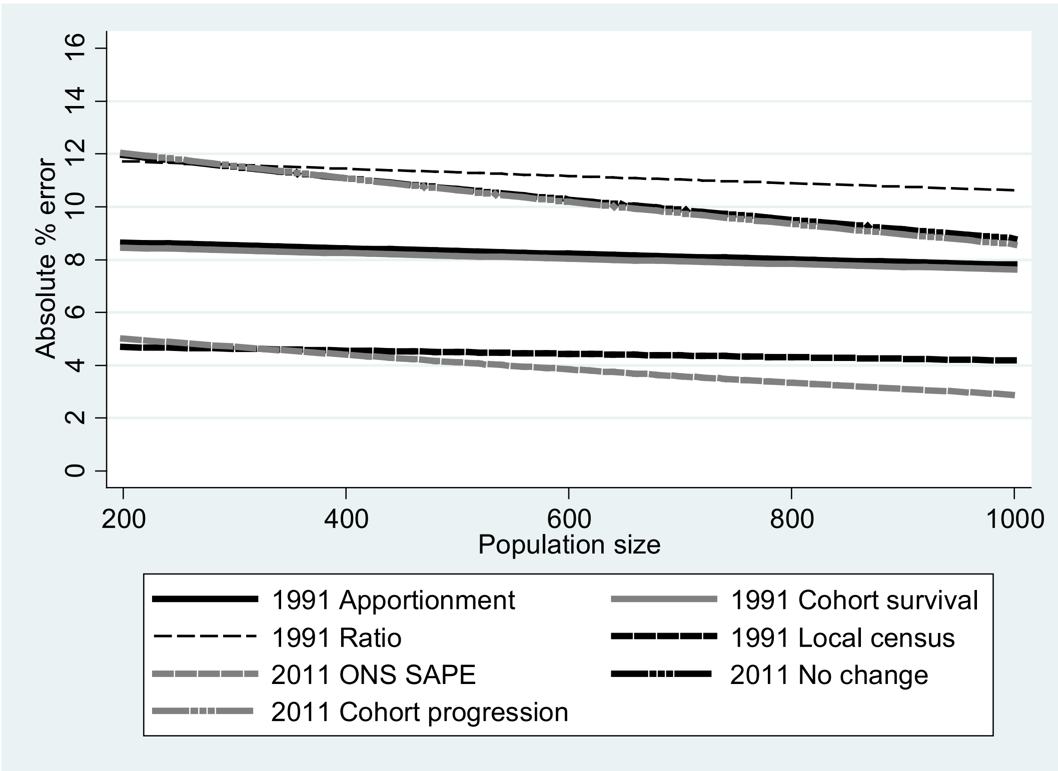
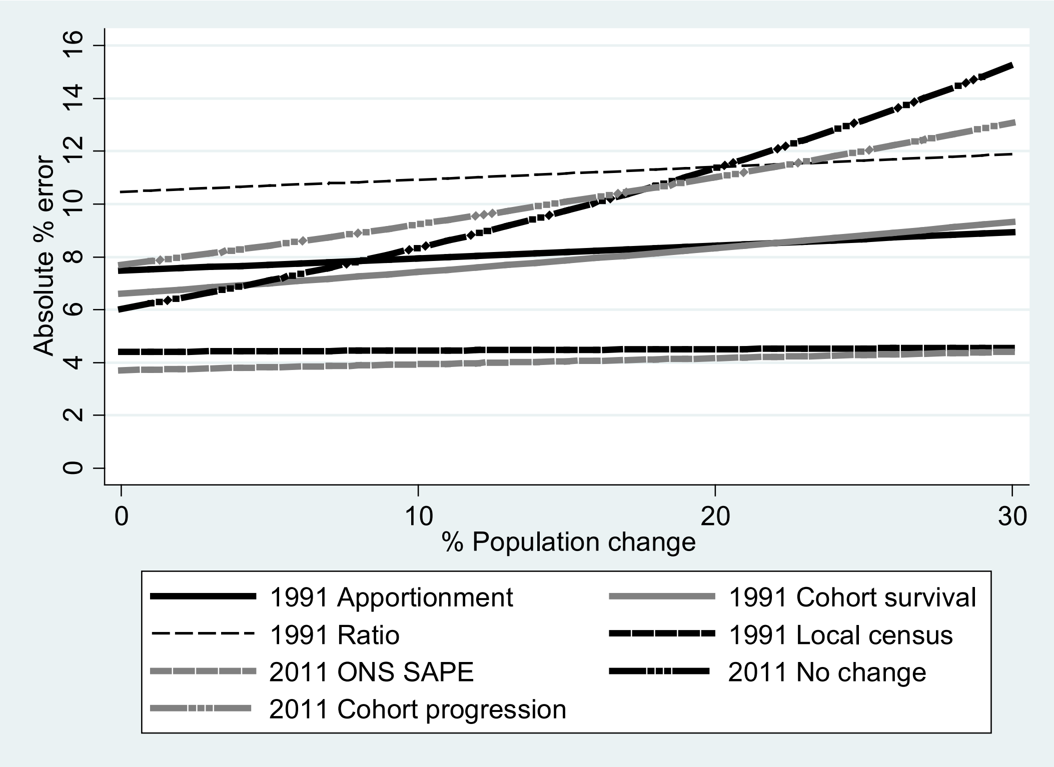


Figure 4 continued overleaf

4c. Model absolute % error by % population change and method (1991and 2011)



4d. Model absolute % error by % in-migration and method (1991and 2011)

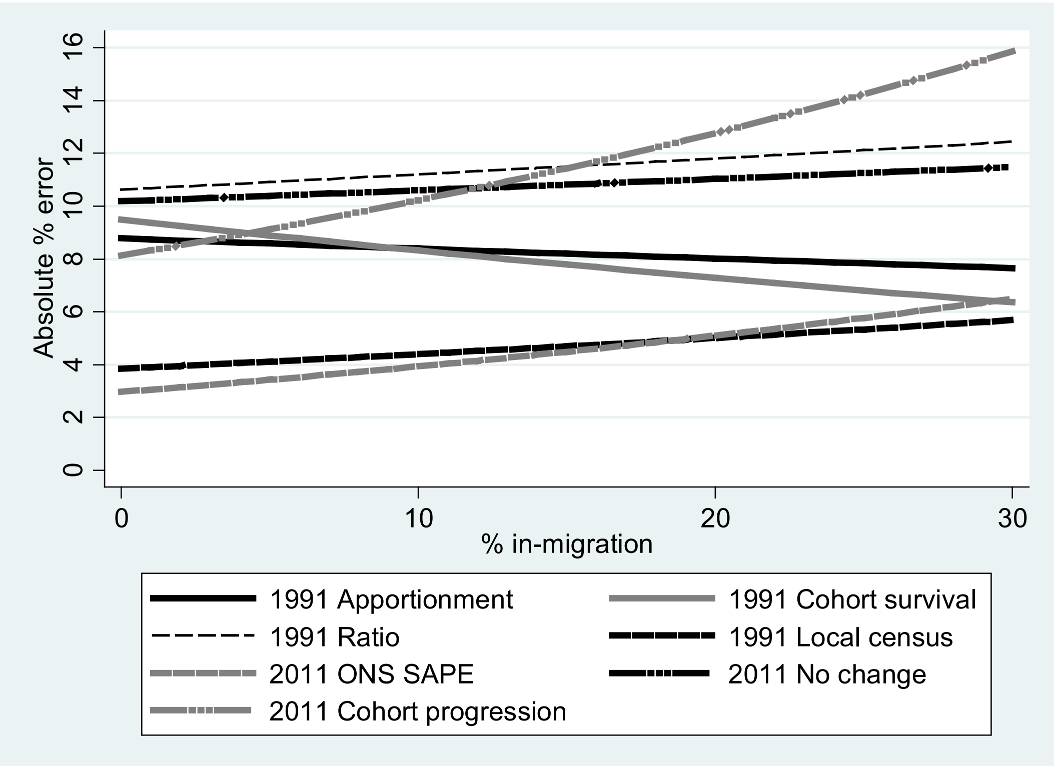
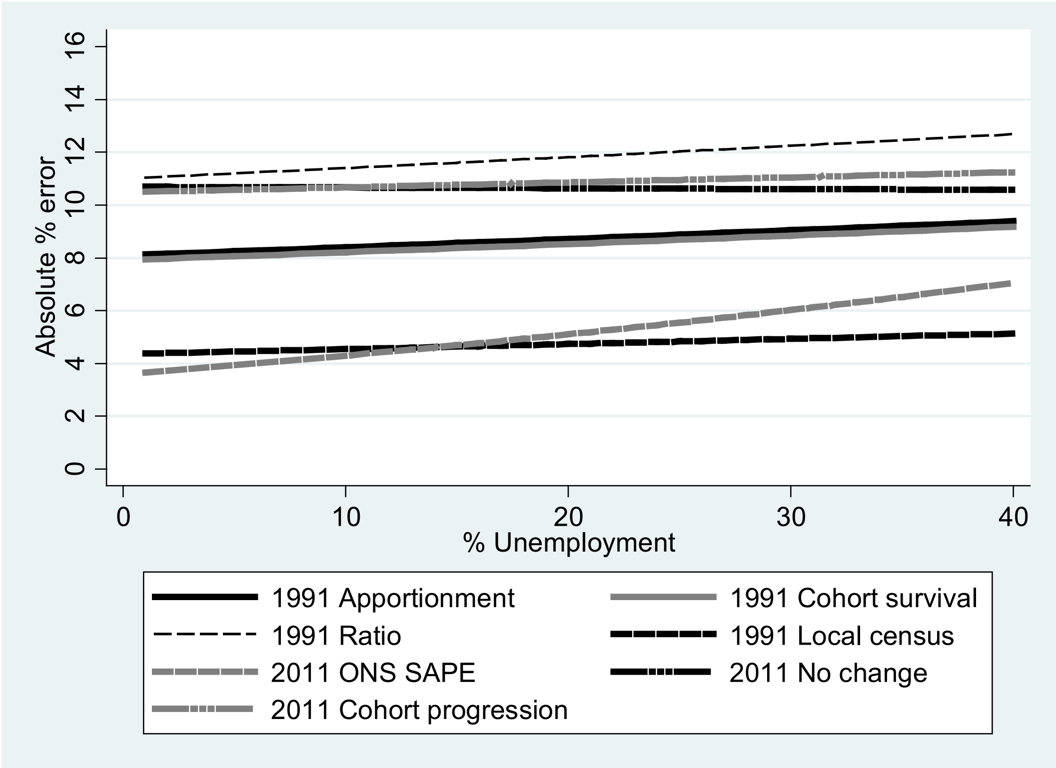


Figure 4 continued overleaf

4e. Model absolute % error by % unemployed and method (1991and 2011)



4f. Model absolute % error by % institutional population and method (1991and 2011)

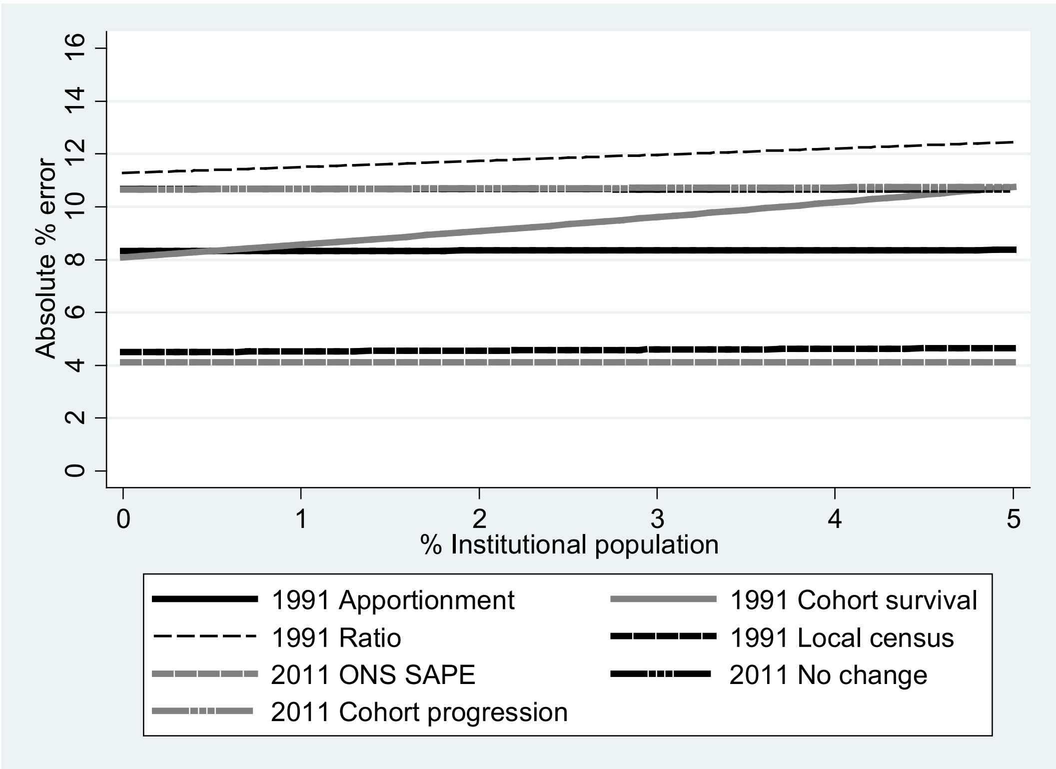


Figure 4: Model absolute % error according to area characteristics. Model estimates relate to the mean (2011) value of the area characteristic not displayed in the graph (see table 1), a population size of 500 and the 0-15 age group.

Tables

Method	Year	Producer	Description
ONS SAPE	2011	Office for National Statistics	Based on ratio change in administrative data then applied (multiplicatively) to a base population. Calibrated to the ONS estimate for the local authority district.
No change	2011	Authors own calculations	Assumes no change to LSOA population. Calibrated to the ONS estimate for the local authority district.
Cohort progression	2011	Authors own calculations	Ages the 2001 census population count at each age group by ten years to derive population estimate in 2011. Calibrated to the ONS estimate for the local authority district.
Local census	1991	Various local planners	Local population estimates generated through additional information requested when updating electoral roll
Cohort survival	1991	Various local planners	Each resident in each small area is aged by the appropriate number of years after accounting for births and deaths and the extent of migration
Apportionment	1991	Various local planners	Indicators of population stock (e.g. electoral register) are used to apportion an independent estimate of a district population to smaller areas within that district.
Ratio	1991	Various local planners	Indicators of population change are used (multiplicatively) to update earlier population estimates for each small area of interest. This method is used in the ONS SAPE.

Table 1: Methods of deriving small area population estimates that are compared in this paper. Note all small area estimates are further calibrated so that age specific LSOA population counts sum to the 'truth' at district level.

Age group	ONS SAPE		No change		Cohort progression	
	Error before calibration	Error after calibration	Error before calibration	Error after calibration	Error before calibration	Error after calibration
0-14	7.4	6.9	17.4	17.4	17.4	17.3
15-29	12.3	9.9	19.1	17.3	24.9	23.1
30-44	8.3	7.0	17.2	17.2	18.1	17.8
45-64	5.3	4.9	12.4	12.2	11.0	10.9
65+	7.2	6.7	17.2	16.9	19.5	19.1

Table 2: Mean absolute % error before and after calibration of LSOA (age-specific) totals to the district population 'truth' (census adjusted mid-year estimate)

Area variables 1991 and 2011	Mean (S.D)				
2011, Unemployment (%)	7.61 (4.72)				
1991, Unemployment (%)	9.97 (5.69)				
2011, Institutional population (%)	1.54 (4.76)				
2011, In-migration 2011 (%)	11.71 (7.36)				
1991, students (%)	2.33 (2.61)				
1991, armed forces (%)	0.04 (0.20)				
Area-age variables 1991 and 2011	**0 to 14**	**15 to 29**	**30 to 44**	**45 to 64**	**65+**
2011, Population size	285 (97.19)	322 (194.70)	333 (106.91)	410 (101.04)	264 (116.12)
1991, Population size	1,614 (926)	1,262 (758)	2,715 (1430)	1,735 (910)	1,250 (721)
2001-2011, Population growth (%)	18.27 (16.1)	16.86 (15.73)	19.72 (16.11)	15.84 (12.60)	19.07 (20.71)
1981-1991, Population growth (%)	14.58 (28.82	15.67 (24.00	21.47 (37.03)	14.99 (23.05	17.25 (27.85
1991, Institutional population (%)[2]	0.50 (1.56)	2.17 (5.32	0.92 (2.14)	0.76 (1.66)	3.33 (3.82
1991, In migration (%)[2]	10.12 (3.86	22.54 (7.87)	10.88 (4.02)	4.35 (2.13	3.33 (1.80)
Dependent variables 1991 and 2011[1]	**0 to 14**	**15 to 29**	**30 to 44**	**45 to 64**	**65+**
2011, ONS SAPE	6.95 (7.47)	9.93 (10.59)	6.96 (6.44)	4.86 (5.40)	6.74 (10.86)
2011, No change	17.39 (15.78)	17.31 (16.99)	17.18 (15.05)	12.20 (12.01)	16.93 (20.18)
2011, Cohort progression	17.39 (15.78)	23.11 (20.40)	17.81 (35.60)	10.85 (13.23)	19.07 (26.76)
1991, Apportionment	9.65 (8.73)	11.37 (10.34)	7.94 (6.96)	9.11 (8.86)	11.70 (10.77)
1991, Cohort survival	7.37 (6.12)	13.29(12.52)	4.53 (7.24)	4.85 (7.98)	6.78 (7.29)
1991, Ratio change	11.99 (9.62)	14.37 (13.03)	7.77 (6.99)	6.03 (5.33)	9.23 (7.75)
1991, Local census	5.97 (10.03)	8.27 (12.10)	4.81 (11.28)	3.86 (9.17)	5.93 (18.64)

Table 3: Mean values (with standard deviation in brackets) of area and area-age variables in 1991 and 2011

1 The dependent variables are described in the text as E, absolute percentage error in small areas calibrated to the District 'truth' in the relevant census year, before log transformation. In 1991, the age groups were the same as shown except for the two younger adult groups which were 15-24 and 25-44.

2 These variables were recorded with age detail in 1991 but not in 2011 where they apply to the total population. In 2011 we base our analysis on LSOAs which have a generally smaller population size that did not support the inclusion of age detail for the variables in question.

Model parameters		Model 1 (Variance component) Coeff	Std error	P<|z|	Model 2 Coeff	Std error	P<|z|	Model 3 Coeff	Std error	P<|z|
Constant term		2.3084	0.0020	<0.0001	1.8449	0.0042	<0.0001	1.8447	0.0041	<0.0001
Age (reference=0-14)										
Age 15 to 29					0.2897	0.0057	<0.0001	0.2907	0.0057	<0.0001
Age 30 to 44					0.0301	0.0057	<0.0001	0.0291	0.0057	<0.0001
Age 45 to 64					-0.1909	0.0060	<0.0001	-0.1893	0.0059	<0.0001
Age 65+					-0.0959	0.0057	<0.0001	-0.0965	0.0057	<0.0001
Method (reference SAPE method)										
No change					0.7287	0.0057	<0.0001	0.7291	0.0058	<0.0001
Cohort survival					0.7309	0.0057	<0.0001	0.7313	0.0059	<0.0001
Population size (2001)					-0.0005	0.0000	<0.0001	-0.0005	0.0000	<0.0001
Absolute population change (2001-2011)					0.0034	0.0001	<0.0001	0.0040	0.0001	<0.0001
In-migration					0.0191	0.0003	<0.0001	0.0188	0.0003	<0.0001
Unemployment					0.0125	0.0005	<0.0001	0.0126	0.0004	<0.0001
Institutional population					-0.0004	0.0005	0.367	-0.0003	0.0005	0.481
No change	Age 15 to 29				-0.2601	0.0081	<0.0001	-0.2619	0.0080	<0.0001
	Age 30 to 44				-0.0529	0.0081	<0.0001	-0.0512	0.0081	<0.0001
	Age 45 to 64				-0.0051	0.0084	0.545	-0.0084	0.0084	0.319
	Age 65+				0.0335	0.0081	<0.0001	0.0346	0.0080	<0.0001
Apportionment	Age 15 to 29				-0.0107	0.0081	0.185	-0.0128	0.0080	0.111
	Age 30 to 44				-0.1196	0.0081	<0.0001	-0.1174	0.0081	<0.0001
	Age 45 to 64				-0.1436	0.0084	<0.0001	-0.1473	0.0084	<0.0001
	Age 65+				0.1141	0.0081	<0.0001	0.1153	0.0080	<0.0001
No change	Population size (2001)				0.0001	0.0000	<0.0001	0.0001	<0.0001	<0.0001
Cohort survival	Population size (2001)				0.0001	0.0000	<0.0001	0.0001	<0.0001	<0.0001
No change	Absolute population change (2001-2011)				0.0234	0.0002	<0.0001	0.0222	0.0002	<0.0001
Cohort survival	Absolute population change (2001-2011)				0.0126	0.0002	<0.0001	0.0111	0.0002	<0.0001
No change	In-migration				-0.0160	0.0004	<0.0001	-0.0154	0.0004	<0.0001
Cohort survival	In-migration				-0.0002	0.0004	0.613	0.0005	0.0005	0.273
No change	Unemployment				-0.0126	0.0006	<0.0001	-0.0129	0.0006	<0.0001
Cohort survival	Unemployment				-0.0108	0.0006	<0.0001	-0.0111	0.0006	<0.0001
No change	Institutional population				0.0004	0.0006	0.53	0.0002	0.0007	0.771
Cohort survival	Institutional population				0.0022	0.0006	<0.0001	0.0020	0.0007	0.005

	Random effects	Estimate	Std. error	95% c.i		Estimate	Std. error	95% c.i		Estimate	Std. error	95% c.i	
Level 1	Variance(Constant)	0.7368	0.0015	0.7338	0.7398	0.5331	0.0011	0.5309	0.5353	0.5266	0.0012	0.5243	0.5289
Level 2 (area)	Variance(Constant)	0.0825	0.0010	0.0805	0.0845	0.0320	0.0005	0.0310	0.0331	0.0271	0.0011	0.0250	0.0292
Level 2 (area)	Covariance(Constant, No change)									-0.0165	0.0012	-0.0189	-0.0142
Level 2 (area)	Variance(No change)									0.0324	0.0020	0.0286	0.0362
Level 2 (area)	Covariance(Constant, Cohort survival)									-0.0158	0.0013	-0.0183	-0.0133
Level 2 (area)	Covariance(No change, cohort survival)									0.0725	0.0018	0.0691	0.0759
Level 2 (area)	Variance(Cohort survival)									0.0675	0.0022	0.0632	0.0719

Table 4: Multilevel model regression estimates

Unemployment	In-migration	10th percentile Pop=183	25th percentile Pop=237	Median Pop=305	75th percentile Pop=390	90th percentile Pop=475
10th percentile (3.3%)	10th percentile (6.3%)	4.5	4.4	4.3	4.1	3.9
	Lower quartile (7.6%)	4.7	4.5	4.4	4.3	4.1
	Median (9.6%)	4.9	4.8	4.7	4.5	4.3
	Upper quartile (13.1%)	5.4	5.3	5.1	4.9	4.8
	90th percentile (19.2%)	6.3	6.2	6.0	5.8	5.6
Lower quartile (4.20%)	10th percentile (6.3%)	4.6	4.5	4.3	4.2	4.0
	Lower quartile (7.6%)	4.7	4.6	4.5	4.3	4.2
	Median (9.6%)	5.0	4.9	4.8	4.6	4.4
	Upper quartile (13.1%)	5.5	5.4	5.2	5.0	4.9
	90th percentile (19.2%)	6.4	6.3	6.1	5.9	5.7
Median (6.00%)	10th percentile (6.3%)	4.7	4.6	4.5	4.3	4.2
	Lower quartile (7.6%)	4.9	4.8	4.7	4.5	4.3
	Median (9.6%)	5.2	5.1	4.9	4.8	4.6
	Upper quartile (13.1%)	5.7	5.5	5.4	5.2	5.0
	90th percentile (19.2%)	6.6	6.5	6.3	6.1	5.9
Upper quartile (9.70%)	10th percentile (6.3%)	5.1	5.0	4.8	4.7	4.5
	Lower quartile (7.6%)	5.3	5.2	5.0	4.8	4.7
	Median (9.6%)	5.6	5.4	5.3	5.1	4.9
	Upper quartile (13.1%)	6.1	5.9	5.8	5.6	5.4
	90th percentile (19.2%)	7.1	6.9	6.7	6.5	6.3
90th Percentile (14.40%)	10th percentile (6.3%)	5.6	5.4	5.3	5.1	4.9
	Lower quartile (7.6%)	5.8	5.6	5.5	5.3	5.1
	Median (9.6%)	6.1	5.9	5.8	5.6	5.4
	Upper quartile (13.1%)	6.6	6.5	6.3	6.1	5.9
	90th percentile (19.2%)	7.7	7.5	7.3	7.1	6.9

Table 5. Expected errors of ONS SAPE for varying population size, unemployment and migration

Note: Errors compare the ONS SAPE without adjustment to the District values of the 2011 Census, to the Census-based mid-2011 population estimate. The values are based on a model with the same transformation and same independent variables as Model 3 in Table 4, and are for the age group 0-14, for an area with mean value on other variables taken from Table 3. Estimates of regression parameters are available on request from the author.